LOCATING THE BUS LOGISTIC NETWORK POINTS TO IMPROVE OPERATIONAL CAPABILITY AT COIMBATORE, INDIA.

- **A work by SRIMAT M**

IBM applied data science capstone

Week5 – final report on capstone project (June 2021)

Table Of Contents

1.INTRODUCTION

The Coimbatore district is a mid-urban district of state Tamilnadu in India. This place called as Manchester of south India and have lots of geographic feature to help the coordination of the logistic in southern india. the geographic position of this district and surrounding region connects the 3 major states of south india (tamilnadu, kerala and Karnataka). also growing urbanization of location makes is budding hotspot for industrial expectations. hence improved logistics will tremendously support the effort for further growth of entire society. Coimbatore is the industrial heart of southern india. Major raw material, composite manufacturing industries and textile industries are around the locality and in demand for the daily logistics all over the india.

How to locate, where to locate the position for logistic points shall be decided by the analysis.

Hence optimal positioning of logistic points will help the zone to improve its transportation demands of locality.

2.TARGET AUDIENCE

The target audience of this project is

- public administrative sectors
- logistics based startup companies
- supply chain and logistic industries

  anyone who is interested in improvising logistics in the locality of Coimbatore surrounding, who are struggling in positioning the perfect location for the connecting hub.

3.DATA COLLECTION

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

• The geographical coordinates of the neighbourhoods are collected using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

• The details for the mall location is scrapped from the FourSquare developers website.

the data was sourced from various websites, csv files and government datas in repositories. and then cleaned and combined for optimal use.

data used.

1. the data provided by the indiapost organisation (in csv form)
2. the data provided by the government of india (https://data.gov.in/catalog/all-india-pincode-directory?filters%5Bfield_catalog_reference%5D=85840&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc)
3. the geolocation data provide by indiapost and kaggle.com
4. webscrapped data from ("https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Coimbatore")

hence all the data is combined with the commonality in single feature called as pincode/postalcode.

hence india is diversified nation and also higly populated every region has lots of subcategory and neighborhood. hence it is mandatory to get deep down and link all the regions.

The following attributes are collected from this data sources,

- Neighborhood(3 levels of sub categories in region)

- Latitude

- Longitude

- Venue Name

- Venue Latitude

- Venue Longitude

- Venue Category

4.0 METHODOLOGY

 For the Capstone project I have analyze and select the best locations for setting up the logistics connectivity hub to improve transportation stability in Coimbatore region.

This being a Data science project the methodology flow as,

 1. Understanding the business problem

2. Data collection

3. Data cleaning

4. Modelling

5. Data Visualization

6. Business insights(analysis and interpretaions)

7. Data Driven decision making

Make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "logistic connectivity" data, we will filter the "bus and transportation connection points" as venue category for the neighbourhoods. Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for "Shopping Mall". The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

5. RESULTS:

The results from the k-means clustering show that we categorize the neighborhoods into 3 Clusters are based on the frequency of occurrence for "bus and transport stations":

➢ Cluster 0 [Red]: Areas with high number of logistic points

➢ Cluster 1 [Purple]: Areas with moderate number to no existence of logistic points

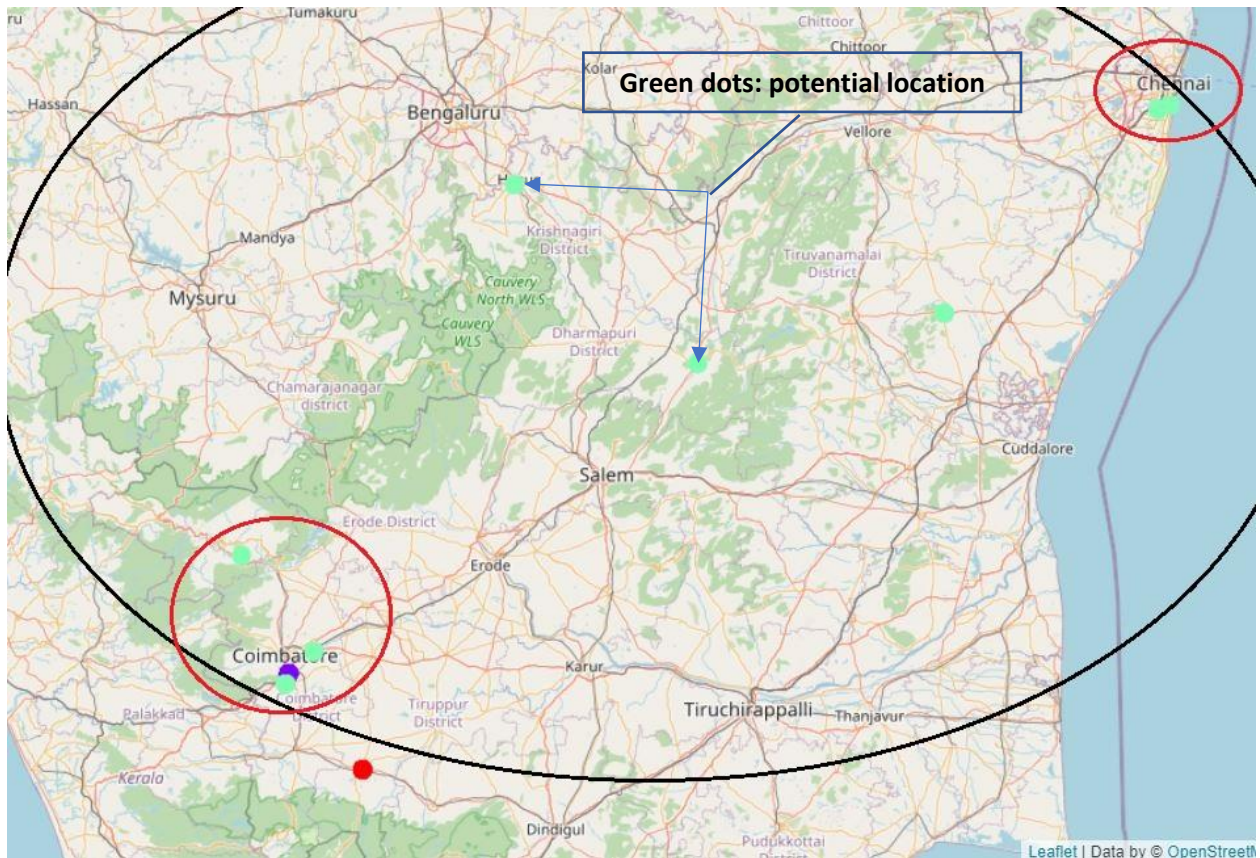➢ Cluster 2 [Green]: Areas with low/ insufficient number of logistic point for operation

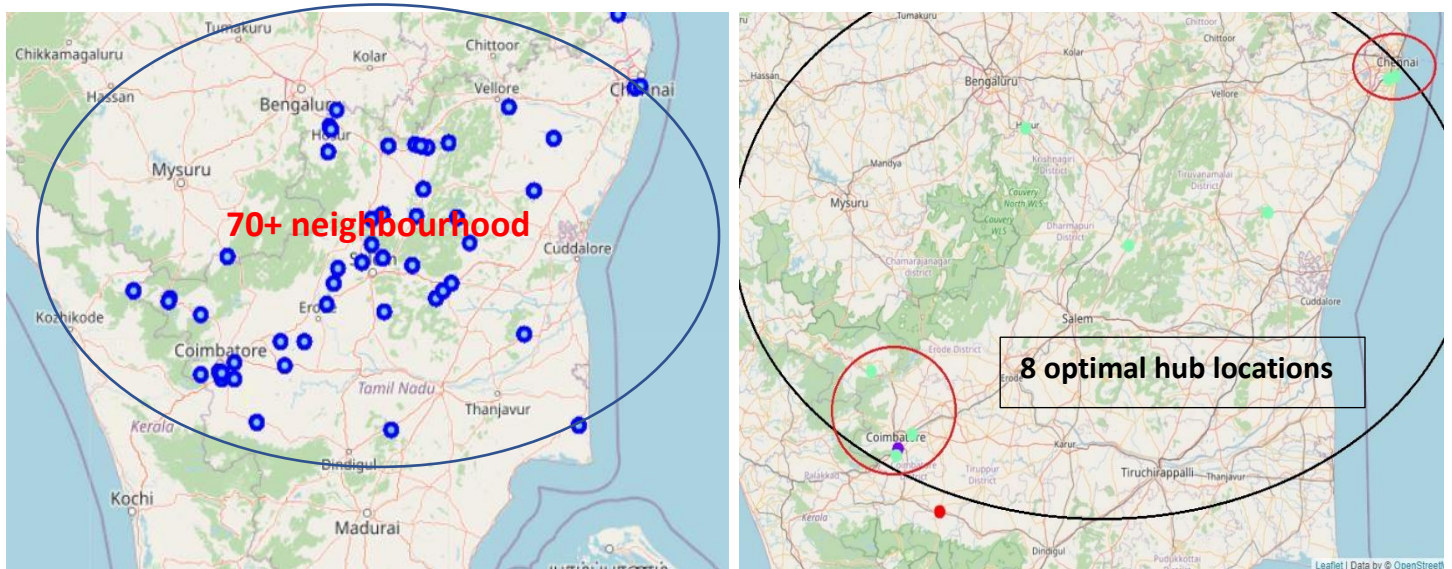Fig.1 . the k mean cluster in search for potential logistic hub placement

6. CONCLUSION:

Fig.2. postions of neighborhood(left) and optimal logistic hub(right)

the conclusion is crucial and significant. the above image are the possible neighborhood of region and the positions to place the potential logistic hub to connect 70+ neighborhood regions.

- Less than 10 logistic hub positions shall connect 70+ neighborhoods in greater way

- These positions shall well connect will possibly increase the overall transport mobility in all three states just by connecting the locations of Coimbatore and Chennai region (red circled in fig.2)

The regions of Chennai and coimbatore(the area covered in red circle marking) are highly demanding zones. As Chennai the tamilnadu capital and port region and in contact with 2 states logistic demand is high. and Coimbatore connecting 3 state also have high demand for logistic connectivity and also middle city in peninsular region of southern india. hence connecting the regions covered by green dots the regions of Coimbatore, erode, Salem to be well connected for the optimal logistic in the region.

7. LIMITATIONS AND FURTHER SCOPE:

- The data from foursquare is not rich, as major population uses google data. And data is not up to date in some locations. Hence accuracy deplete in minor ways.

- Another limitation is analysis purely based on bus transport logistics and road transportation. And excluding the railway connectivity.

The future scope shall be the comparing the neighbor hood of key district in three states including road and rail transport to find the optimal positioning of new logistic hub for effective transportations.

REFERENCES:

- https://data.gov.in/catalog/all-india-pincode-directory?filters%5Bfield_catalog_reference%5D=85840&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc

- https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Coimbatore

- www.kaggle.com

- Documentations of python, beautiful soup and foursquare api